Frozen Lake

班級 | 電子四甲
姓名 | 沈酩閎
學號 | 105360043

程式

說明

- import

```
#%%


import os
import random


import gym
import numpy as np
```

程式
說明

- 初始化環境

```
env = gym.make("FrozenLake8x8-v0")
env.reset()
env.render()
```

- 根據Q-Learning公式進行參數調整

We use Q-learning to learn the expected values of all action-state pairs. (Wikipedia). The update formula is shown below:

$$Q^{new}(s_t, a_t) \leftarrow \underbrace{Q(s_t, a_t)}_{\text{old value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \overbrace{\Big(\underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1}, a)}_{\text{estimate of optimal future value}} - \underbrace{Q(s_t, a_t)}_{\text{old value}} \Big)}^{\text{temporal difference}}$$

$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}_{\text{new value (temporal difference target)}}$$

We also adopt ϵ-greedy strategy to explore.

程式
說明

- 參數調整說明

```
actions = env.action_space.n
states = env.observation_space.n
eposides = 100000000 #eposides = 100000
epsilon = 0.99 #epsilon = 0.8
gamma = 0.99 #gamma = 0.9
alpha = 0.005 #alpha = 0.01
```

- eposides = 100000000
  更多的訓練次數，理想是無限多
- gamma = 0.99
  比預設值更趨近於1，更重視長期的獎勵
- alpha = 0.005
  降低學習率，但要提高訓練次數才能達到加強學習的效果
- eposilon = 0.99
  提高Greedy初始搜尋路徑的廣泛性，比較慢收斂，但藉由訓練次數增加會有較好的學習效果

程式
說明

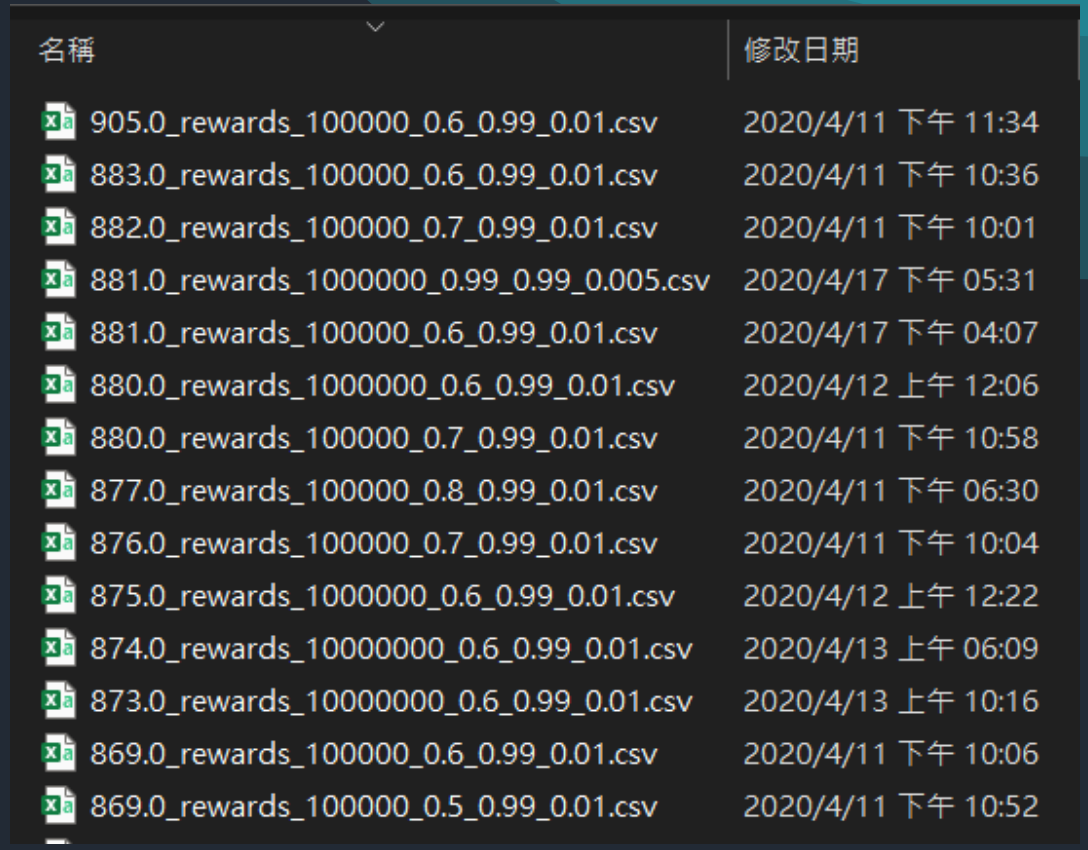
- 設定輸出資料夾及CSV檔名稱

```
filename = 'rewards_%s_%s_%s_%s' %(eposides, epsilon, gamma, alpha)
outputDir = './output'
```

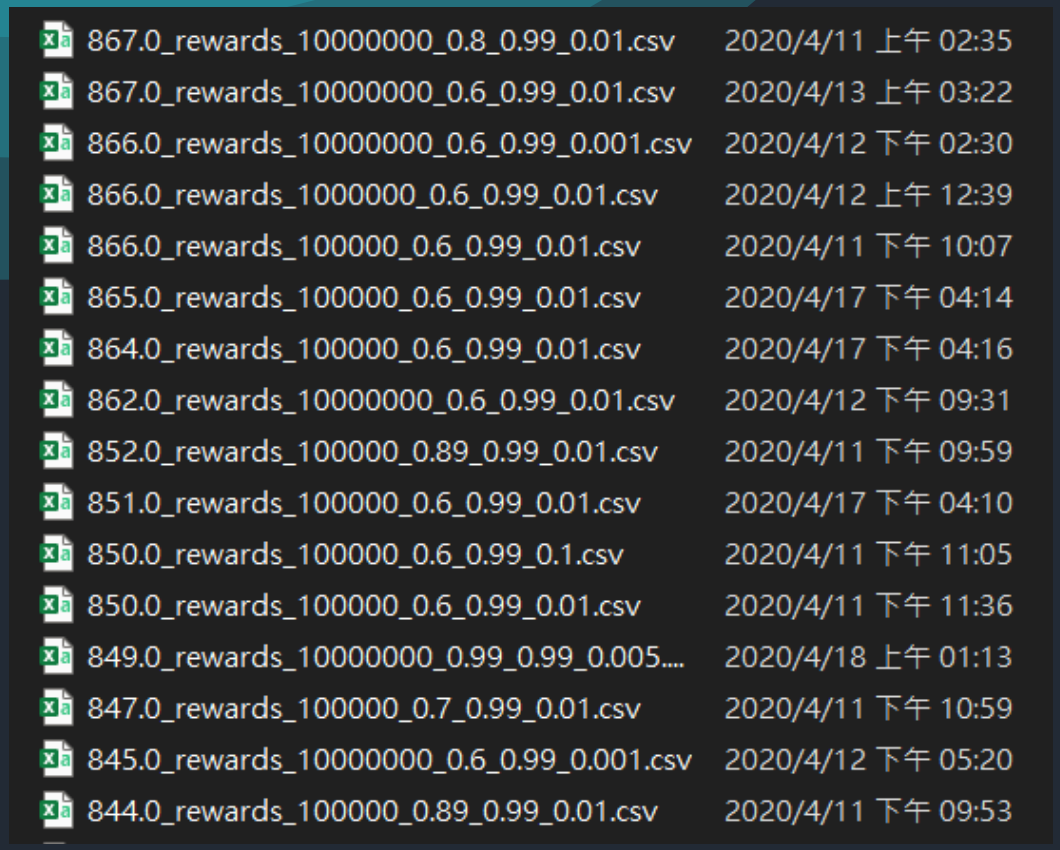
- 輸出資料夾及CSV檔

```
filename = str(total_reward) + '_' + filename + '.csv'
total_avg_reward = total_reward/test_episodes
# Print results in CSV format and upload to Kaggle
if not os.path.exists(outputDir):
    os.mkdir(outputDir)
with open('./output/%s' %filename, 'w') as f:
    f.write('Id,Predicted\n')
    f.write('FrozenLake8x8_public,{}\n'.format(total_avg_reward))
    f.write('FrozenLake8x8_private,{}\n'.format(total_avg_reward))
```

測試結果



名稱	修改日期
905.0_rewards_100000_0.6_0.99_0.01.csv	2020/4/11 下午 11:34
883.0_rewards_100000_0.6_0.99_0.01.csv	2020/4/11 下午 10:36
882.0_rewards_100000_0.7_0.99_0.01.csv	2020/4/11 下午 10:01
881.0_rewards_1000000_0.99_0.99_0.005.csv	2020/4/17 下午 05:31
881.0_rewards_100000_0.6_0.99_0.01.csv	2020/4/17 下午 04:07
880.0_rewards_1000000_0.6_0.99_0.01.csv	2020/4/12 上午 12:06
880.0_rewards_100000_0.7_0.99_0.01.csv	2020/4/11 下午 10:58
877.0_rewards_100000_0.8_0.99_0.01.csv	2020/4/11 下午 06:30
876.0_rewards_100000_0.7_0.99_0.01.csv	2020/4/11 下午 10:04
875.0_rewards_1000000_0.6_0.99_0.01.csv	2020/4/12 上午 12:22
874.0_rewards_10000000_0.6_0.99_0.01.csv	2020/4/13 上午 06:09
873.0_rewards_10000000_0.6_0.99_0.01.csv	2020/4/13 上午 10:16
869.0_rewards_100000_0.6_0.99_0.01.csv	2020/4/11 下午 10:06
869.0_rewards_100000_0.5_0.99_0.01.csv	2020/4/11 下午 10:52



867.0_rewards_10000000_0.8_0.99_0.01.csv	2020/4/11 上午 02:35
867.0_rewards_10000000_0.6_0.99_0.01.csv	2020/4/13 上午 03:22
866.0_rewards_10000000_0.6_0.99_0.001.csv	2020/4/12 下午 02:30
866.0_rewards_1000000_0.6_0.99_0.01.csv	2020/4/12 上午 12:39
866.0_rewards_100000_0.6_0.99_0.01.csv	2020/4/11 下午 10:07
865.0_rewards_100000_0.6_0.99_0.01.csv	2020/4/17 下午 04:14
864.0_rewards_100000_0.6_0.99_0.01.csv	2020/4/17 下午 04:16
862.0_rewards_10000000_0.6_0.99_0.01.csv	2020/4/12 下午 09:31
852.0_rewards_100000_0.89_0.99_0.01.csv	2020/4/11 下午 09:59
851.0_rewards_100000_0.6_0.99_0.01.csv	2020/4/17 下午 04:10
850.0_rewards_100000_0.6_0.99_0.1.csv	2020/4/11 下午 11:05
850.0_rewards_100000_0.6_0.99_0.01.csv	2020/4/11 下午 11:36
849.0_rewards_10000000_0.99_0.99_0.005....	2020/4/18 上午 01:13
847.0_rewards_100000_0.7_0.99_0.01.csv	2020/4/11 下午 10:59
845.0_rewards_10000000_0.6_0.99_0.001.csv	2020/4/12 下午 05:20
844.0_rewards_100000_0.89_0.99_0.01.csv	2020/4/11 下午 09:53

- 檔案命名格式totalReward_eposides_eposilon_gamma_alpha.csv
- 由此可知，調降eposilon至0.6左右可以在小訓練次數看到好的效果，但到達一定訓練次數之後無法繼續進步，必須改為增加eposides、eposides(趨近1)、 gamma (趨近1) 、並隨著訓練次數增加動態減少alpha(由大至小)才能繼續進步下去

Kaggle | Score

Submission and Description	Private Score	Public Score	Use for Final Score
883.0_rewards_100000_0.6_0.99_0.01.csv a day ago by t105360043 add submission details	0.11700	0.11700	☐

18	—	t105360043		0.11700	4	11h

THANK YOU